Empowering Citizens. Smarter Societies.

Insight
Centre for Data Analytics

# Rich-Context: An Unsupervised Context-Driven Recommender System Based On User Reviews


University College Cork

Francisco J. Peña

Supervisor: Dr. Derek Bridge


A World Leading SFI Research Centre

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD Dublin Science Foundation Ireland sfi For what's next

# Problem


McDonald's
HAMBURGERS
BILLIONS & BILLIONS SERVED
MINI MEALS
4
MEALS UNDER
$3
McDonald's

# Context in Recommender Systems

## Context Aware Recommender Systems

- Most of them predefine context
- Small number of features
- Small number of values

| | | | |
|---|---|---|---|
| **I'm travelling for:** | ◯ Work | ◯ Leisure | |
| **Companion:** | ◯ Solo | ◯ Couple | ◯ Family |

## Open-ended context is very wide

- Context is richer, open-ended
- Birthday, anniversary, parking, accessibility, eat-in vs take away, pet friendly, …

# Goals

## Recommendation model

- Treat context as open-ended
- Unsupervised (not predefined keywords)
- Good performance on sparse datasets

## Evaluation methodology

- Datasets
  - Big (+10 000 records)
  - Sparse
  - Multiple from different domains
  - Publicly available
  - Real-world
- Third-party evaluation tool

# Contributions

Recommendations without predefining context

New way of representing context.

# Contributions

- Recommendations without predefining context
  - New way of representing context.
- Better prediction performance
  - New methodology for offline ranking evaluation.
  - Better performance for brand-new users.

# Contributions

- **Recommendations without predefining context**
  - New way of representing context.
- **Better prediction performance**
  - New methodology for offline ranking evaluation.
  - Better performance for brand-new users.
- **Extract context from reviews in an unsupervised way**
  - Methodology for selecting the best topic modeling algorithm.
  - New metrics to measure context-richness of topic models.
  - Improved methodology to classify reviews.

# Assumptions

## Specific Review

- *"During the summer, we like to take a mini staycation. This year it was extra special as we also got engaged. Our stay at the Biltmore was just fantastic. The service exceptional, the food amazing."*

## Generic Review

- *"Nice hotel, all the amenities you need, great complex of pools."*

# Assumptions

## Specific Review

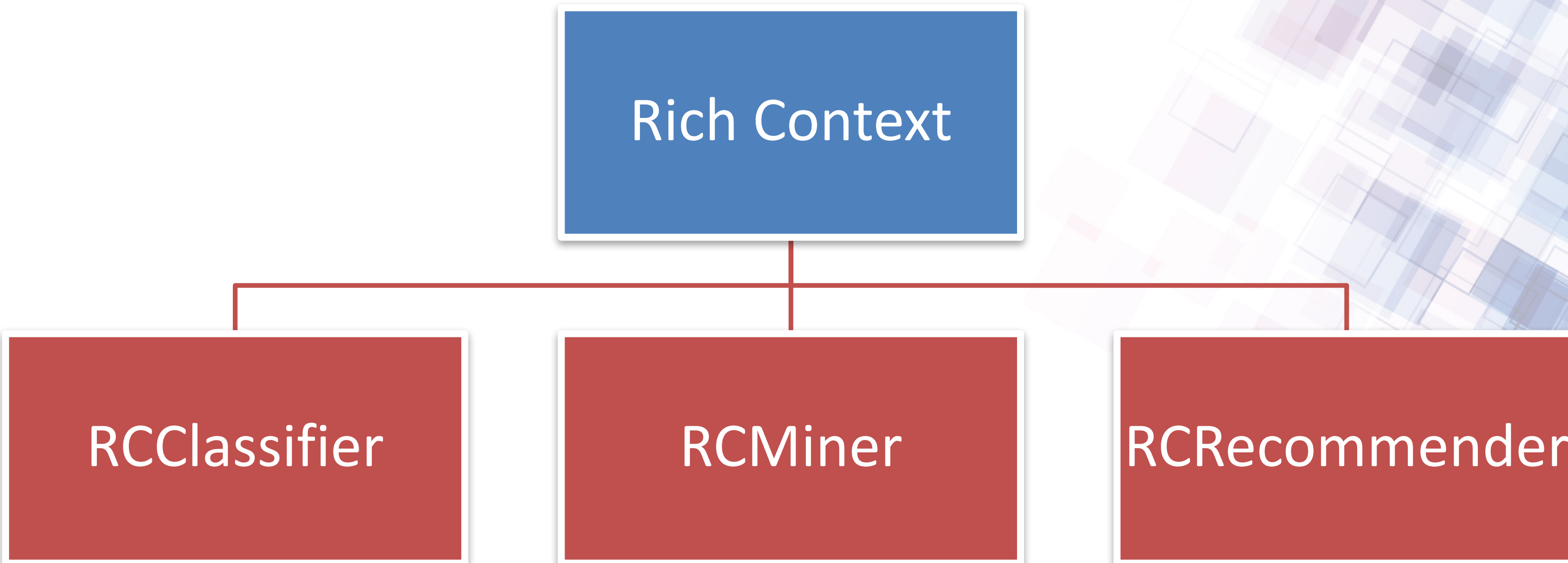
- *“During the summer, we like to take a mini staycation. This year it was extra special as we also got engaged. Our stay at the Biltmore was just fantastic. The service exceptional, the food amazing.”*

Specific reviews contain more contextual information than generic ones.

## Generic Review

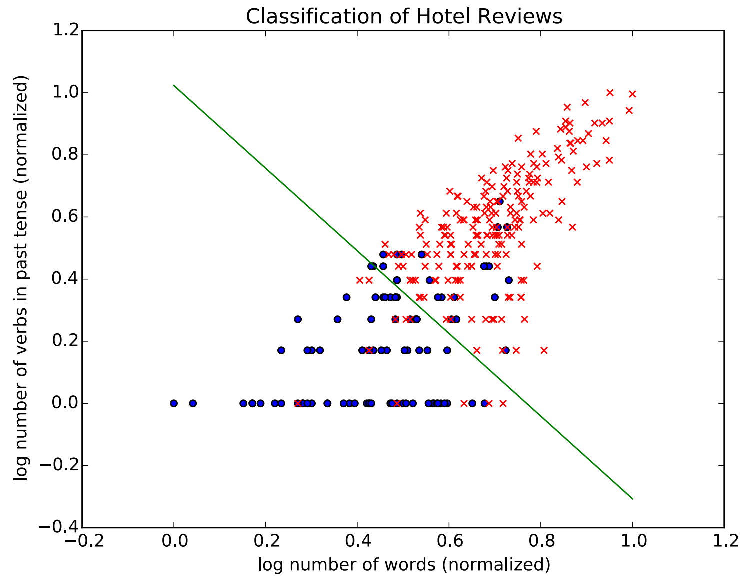
- *“Nice hotel, all the amenities you need, great complex of pools.”*

# Rich Context (RC)

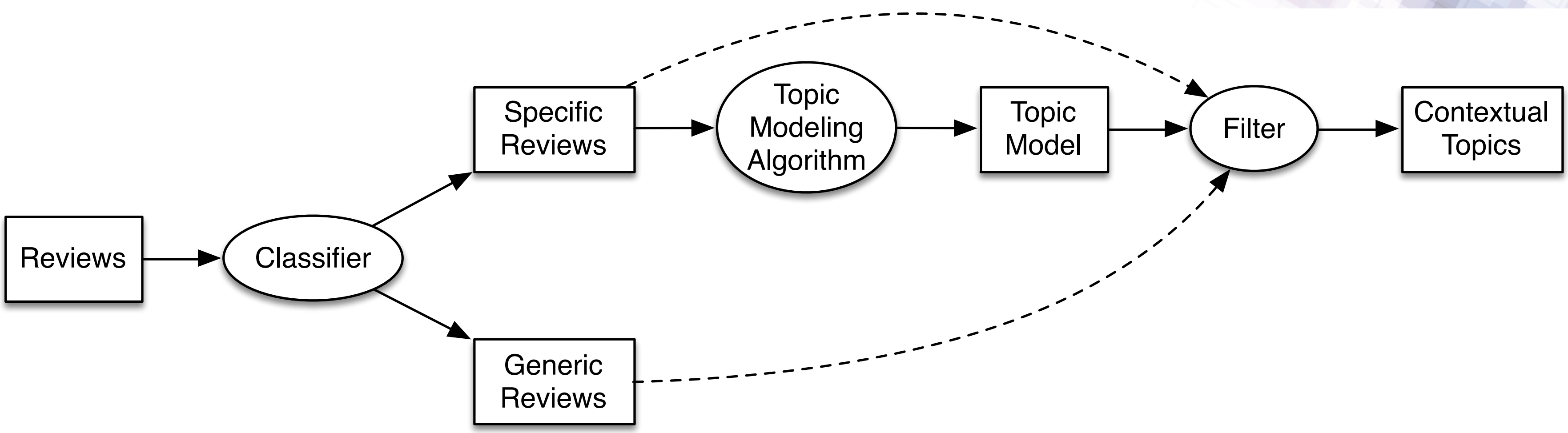


# Reviews Classification

- 300 tagged reviews
- Random Forest Classifier
- Features
  - LogWords: log of number of words in the review + 1
  - Vsum: log of number of verbs in the review + 1
  - VBDSum: log of number of verbs in the past tense in the review + 1
  - ProRatio: ratio of log of number of personal pronouns + 1 to LogWords

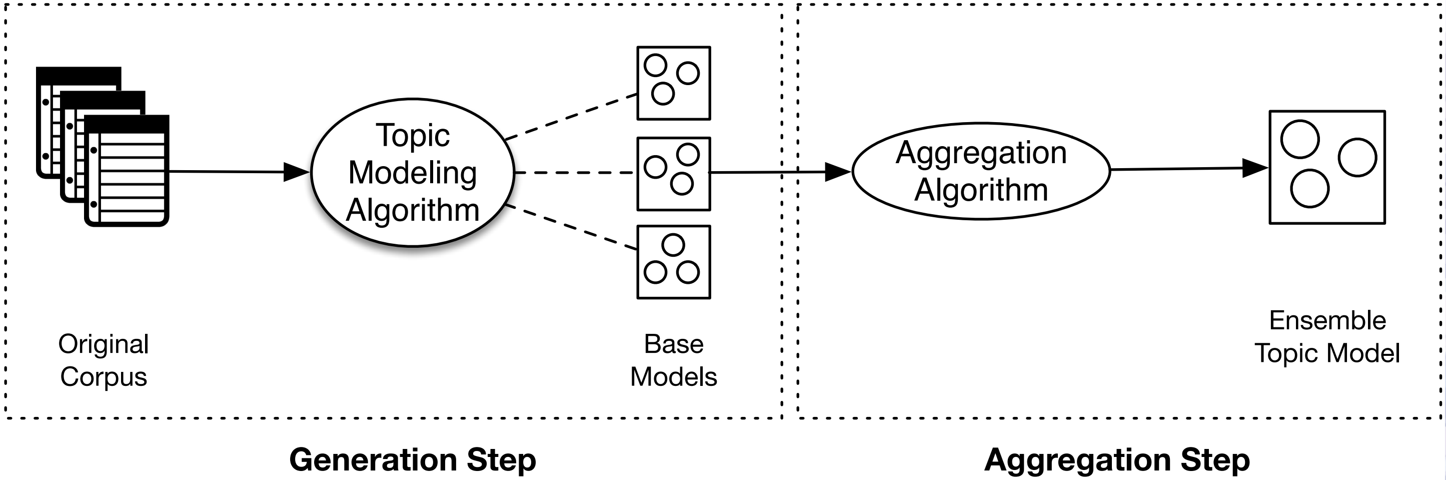


# Context Extraction



# Ensemble Topic Modeling

Source: Stability of Topic Modeling via Matrix Factorization. Belford et al




## Advantages

- More stable topic models.
- Context-richer topics.

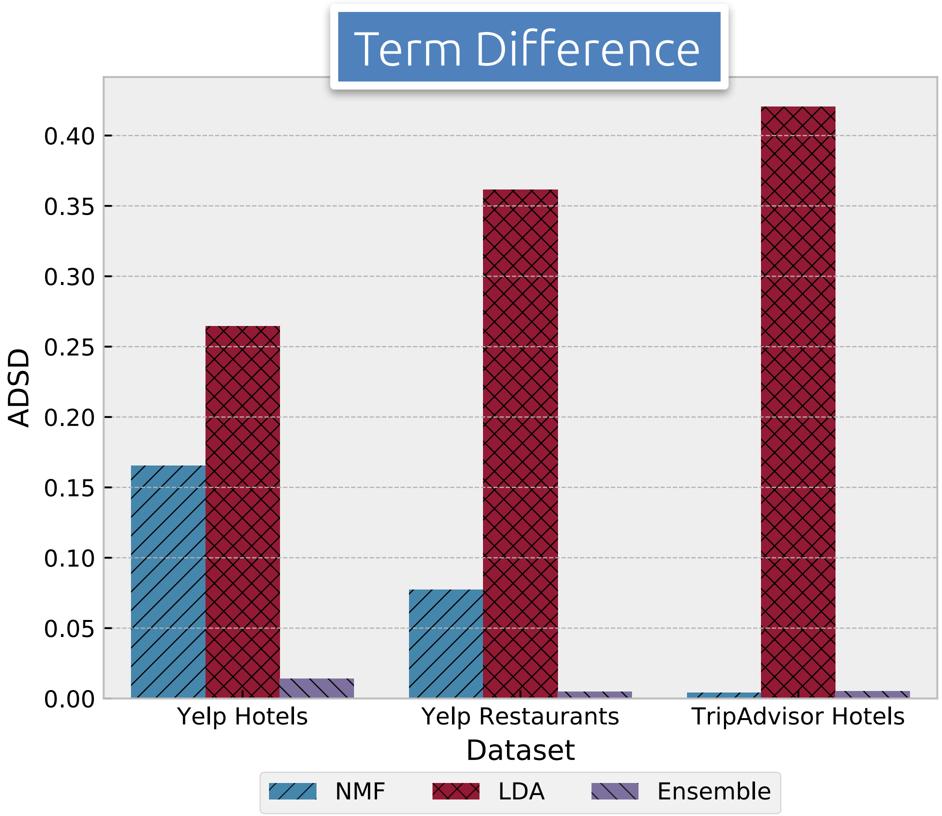
## Topic Model Validation

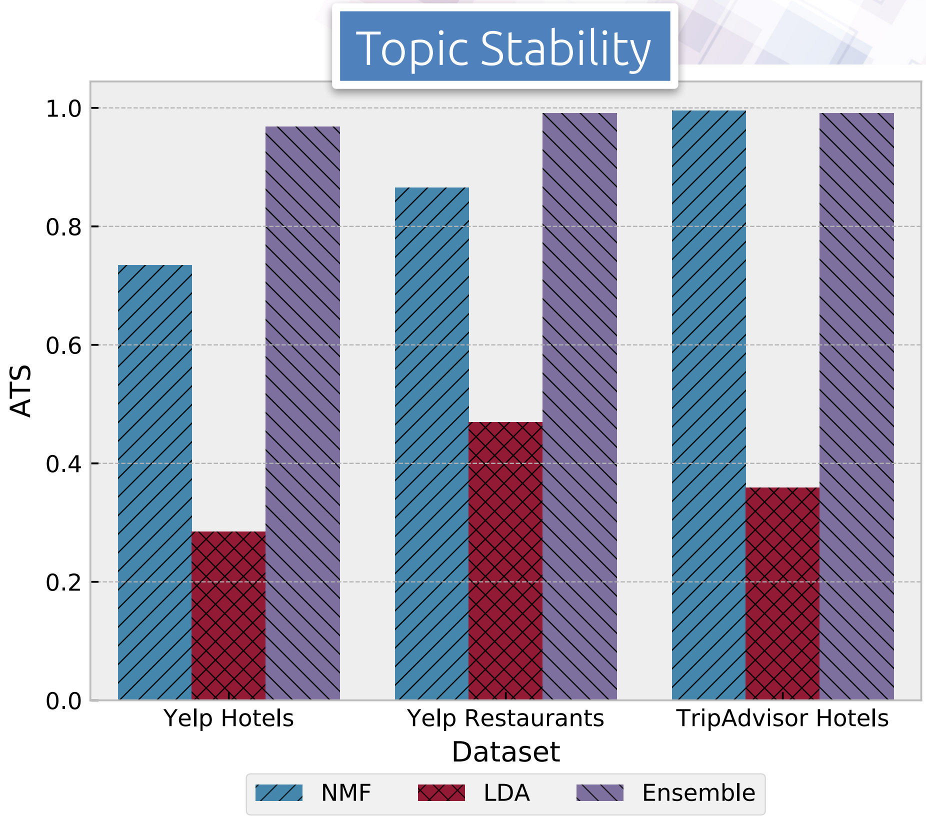
What topic modeling is more stable?

→

Which reviews data produces context-richer topic models?

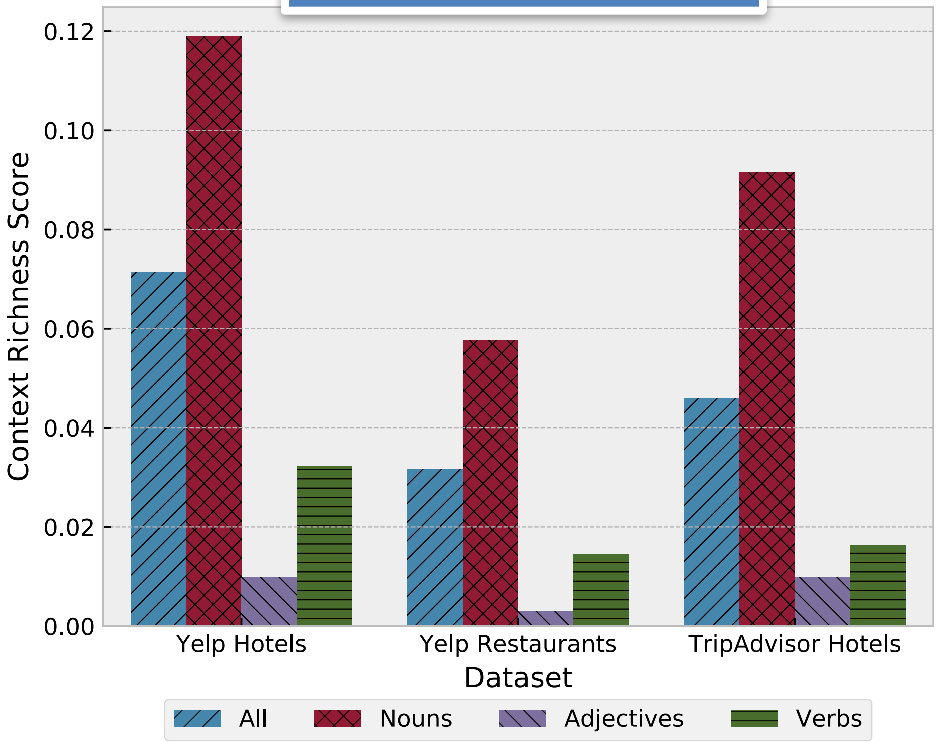
# Topic Model Validation (Stability)



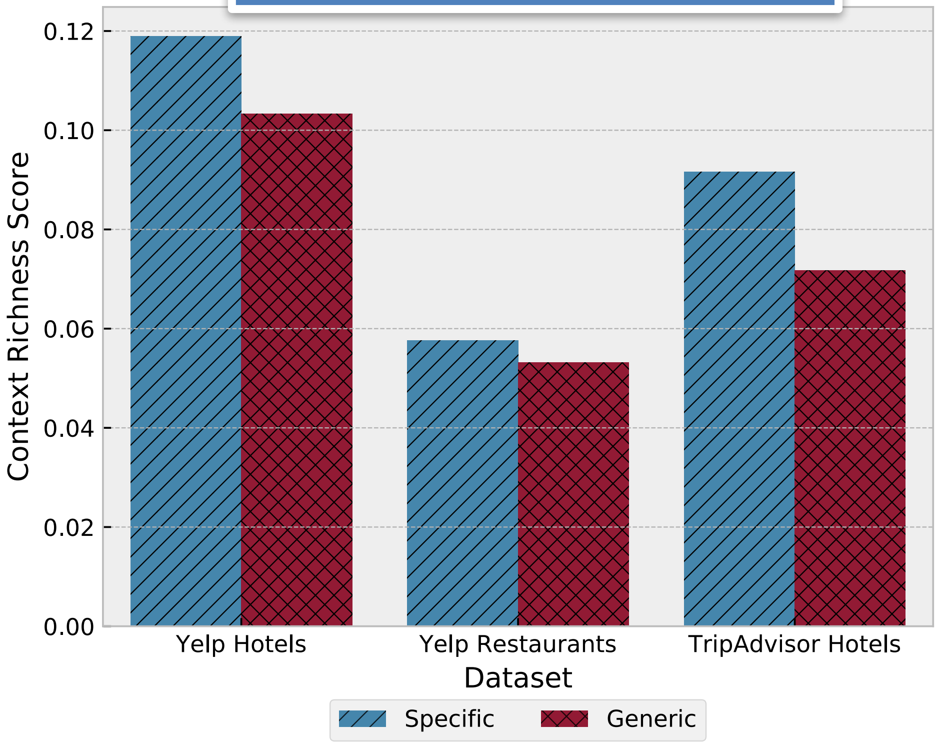



# Topic Model Validation (Context-Richness)

Part of speech types

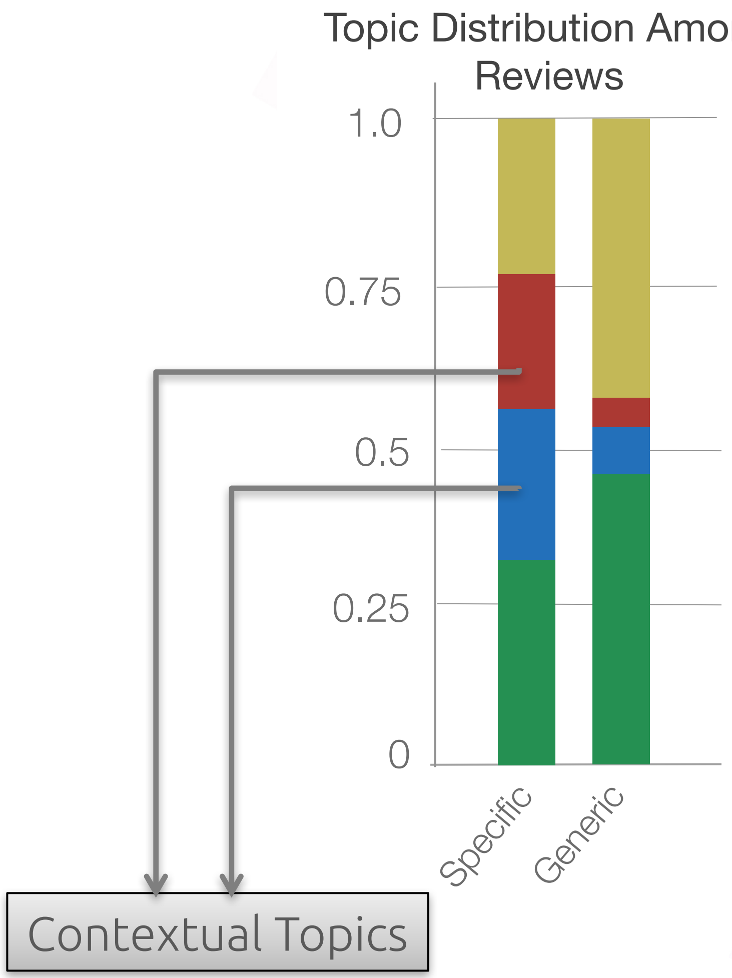



Specific vs generic reviews

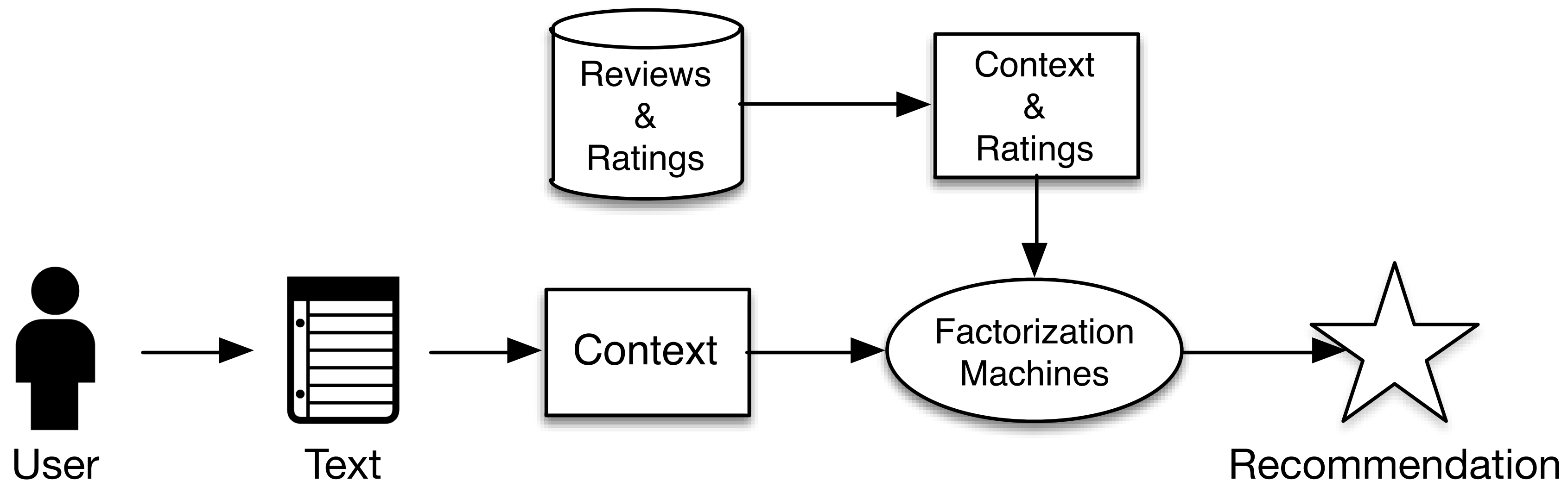


# Context Extraction

- Apply the topic model to both specific and generic reviews.
- Count the number of times topics appear in specific and generic reviews.
- The ones that appear more frequently in specific reviews are labeled as contextual topics.



# Recommendations



# Evaluation - Dataset Description

### Yelp Hotels

- 3,809 reviews
- 3,205 users
- 98 items
- 98.79% sparsity

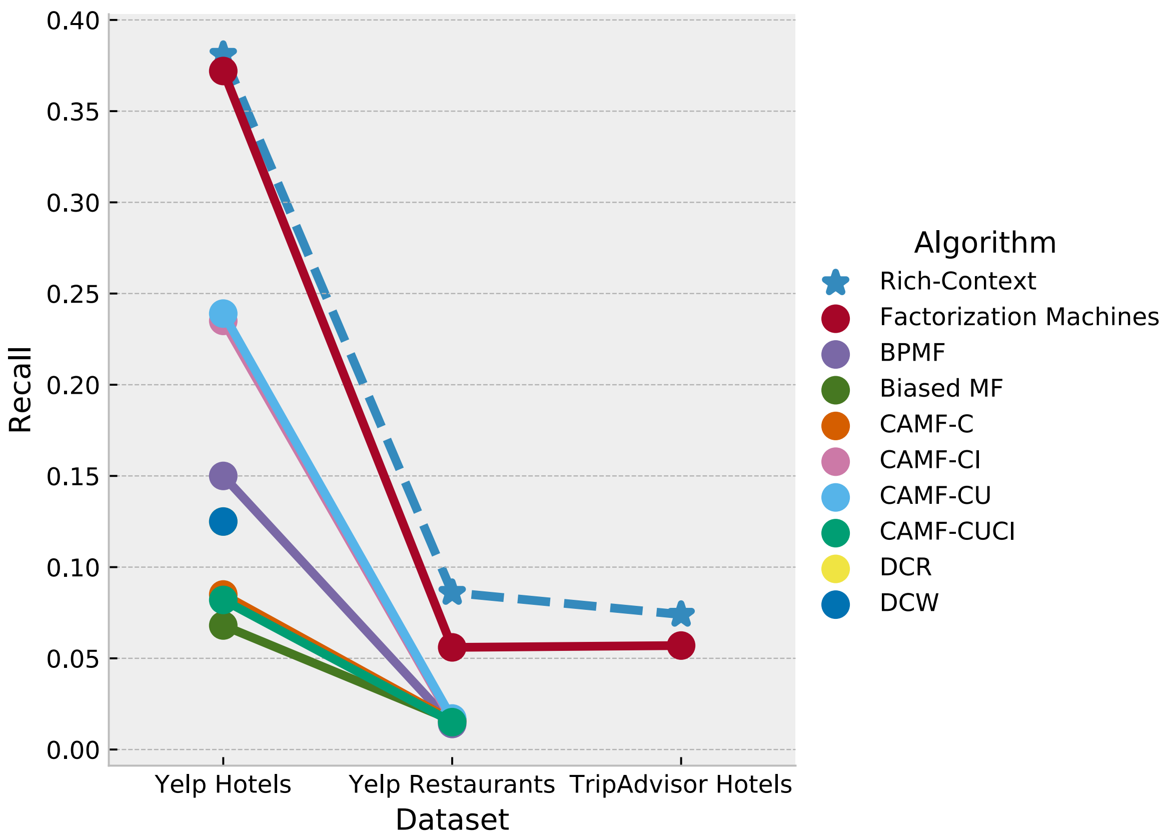
### Yelp Restaurants

- 147,864 reviews
- 35,021 users
- 2,550 items
- 99.83% sparsity

### TripAdvisor Hotels

- 726,426 reviews
- 526,717 users
- 3,299 items
- 99.96% sparsity

# Evaluation - Ranking Prediction (Recall@10)




Compared to best SOTA

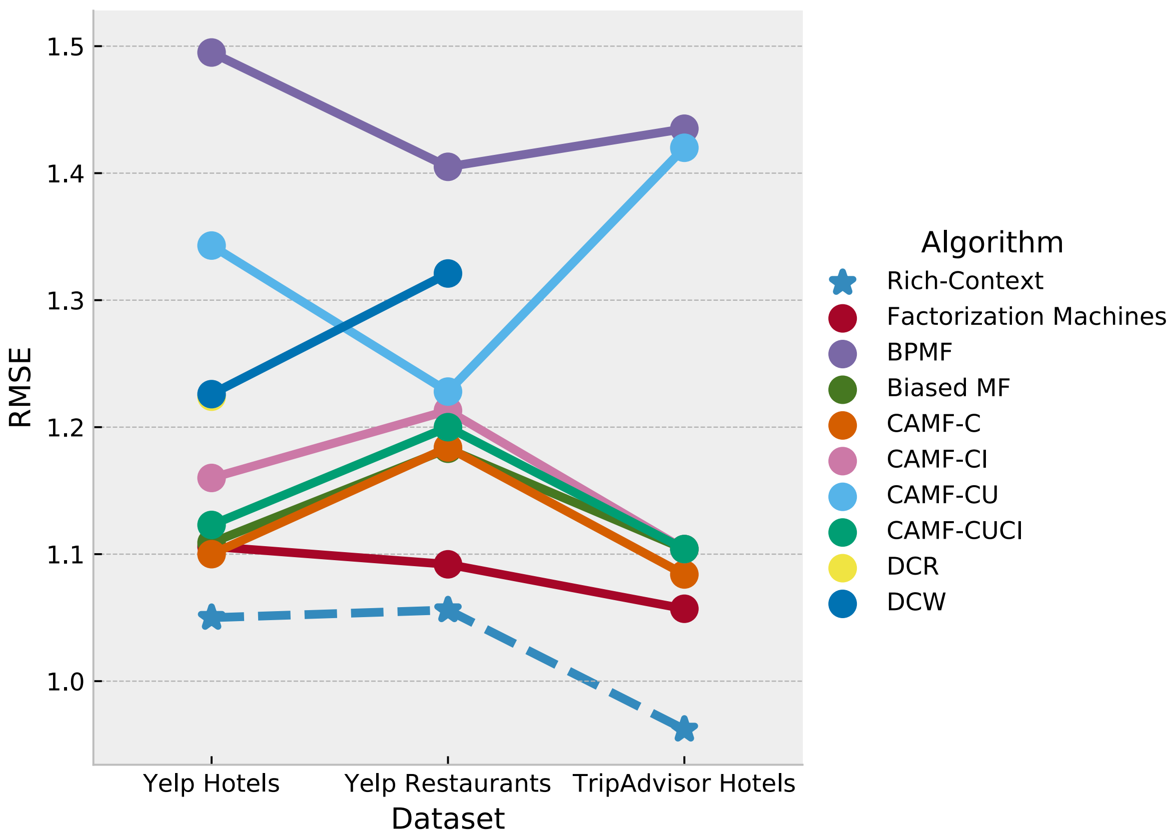
**Improvement (all users)**

- Yelp Hotels (+2.32%)
- Yelp Restaurants (+55.07%)
- TripAdvisor Hotels (+30.15%)

**Improvement (new users)**

- Yelp Hotels (+3.67%)
- Yelp Restaurants (+47.00%)
- TripAdvisor Hotels (+20.50%)

# Evaluation - Rating Prediction (RMSE)




**Compared to best SOTA**

**Improvement (all users)**

- Yelp Hotels (+4.45%)
- Yelp Restaurants (+3.29%)
- TripAdvisor Hotels (+8.99%)

**Improvement (new users)**

- Yelp Hotels (+5.44%)
- Yelp Restaurants (+4.54%)
- TripAdvisor Hotels (+9.96%)

# Final Thoughts

## Conclusions

- We present a context-driven recommender system that does not pre-defined contextual words.
- We improve recommendations by using the mined contextual information as side-information in factorization machines.
- The proposed model does not need expertise about contextual information.

## Future Work

- Improve the topic model quality metrics in order to evaluate topic models without having to run the recommender (like a classifier).
- Use topic models to produce explanations of recommendations.
- Extrapolate the same model to other scenarios where documents are available and recommendations are needed, for instance using medical records, law, etc.

Empowering Citizens. Smarter Societies.

Insight
Centre for Data Analytics

# Thanks!

Acknowledgments
- Derek Bridge
- Diego Carraro
- Mesut Kaya
- Insight UCC
- SFI

francisco.pena@insight-centre.org

A World Leading SFI Research Centre

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD Dublin Science Foundation Ireland sfi For what's next

# References

Adomavicius, G., and Tuzhilin, A. 2011. Context-aware recommender systems. In Ricci, F.; Rokach, L.; Shapira, B.; and Kantor, P. B., eds., *Recommender Systems Handbook*. Springer. 217–253.

Arora, S.; Ge, R.; and Moitra, A. 2012. Learning topic mod- els – going beyond SVD. In *IEEE 53rd Annual Symposium on Foundations of Computer Science*, 1–10.

Baltrunas, L.; Ludwig, B.; and Ricci, F. 2011. Matrix fac- torization techniques for context aware recommendation. In *5th ACM Conference on Recommender Systems*, 301–304.

Batista, G. E. A. P. A.; Prati, R. C.; and Monard, M. C. 2004. A study of the behavior of several methods for balanc- ing machine learning training data. *SIGKDD Explor. Newsl.* 6(1):20–29.

Bauman, K., and Tuzhilin, A. 2014. Discovering contex- tual information from user reviews for recommendation pur- poses. In *1st Workshop on New Trends in Content-based Recommender Systems at the 8th ACM Conference on Rec- ommender Systems*, 2–9.

Belford, M.; Namee, B. M.; and Greene, D. 2016. Ensem- ble topic modeling via matrix factorization. In *24th Irish Conference on Artificial Intelligence and Cognitive Science*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3:993–1022.

Chen, G., and Chen, L. 2015. Augmenting service recom- mender systems by incorporating contextual opinions from user reviews. *User Modeling and User-Adapted Interaction* 25(3):295–329.

# References

Chen, L.; Chen, G.; and Wang, F. 2015. Recommender systems based on user reviews: The state of the art. *User Modeling and User-Adapted Interaction* 25(2):99–154.

Cremonesi, P.; Koren, Y.; and Turrin, R. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *4th ACM Conference on Recommender Systems*, 39–46.

Diao, Q.; Qiu, M.; Wu, C.-Y.; Smola, A. J.; Jiang, J.; and Wang, C. 2014. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In *20th ACM International Conference on Knowledge Discovery and Data Mining*, 193–202.

Greene, D.; O'Callaghan, D.; and Cunningham, P. 2014. How many topics? stability analysis for topic models. In *European Conference on Machine Learning and Knowledge Discovery in Databases*, 498–513.

Hariri, N.; Zheng, Y.; Mobasher, B.; and Burke, R. 2011. Context-aware recommendation based on review mining. In *9th Workshop on Intelligent Techniques for Web Personal- ization & Recommender Systems*, 30–36.

Karatzoglou, A.; Amatriain, X.; Baltrunas, L.; and Oliver, N. 2010. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *4th ACM Conference on Recommender Systems*, 79–86.

Lee, D. D., and Seung, S. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401.

Lemaˆıtre, G.; Nogueira, F.; and Aridas, C. K. 2016. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *CoRR* abs/1609.06570.

# References


Ling, G.; Lyu, M. R.; and King, I. 2014. Ratings meet reviews, a combined approach to recommend. In *8th ACM Conference on Recommender Systems*, 105–112.

McAuley, J., and Leskovec, J. 2013. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *7th ACM Conference on Recommender Systems*, 165–172.

Pagano, R.; Cremonesi, P.; Larson, M.; Hidasi, B.; Tikk, D.; Karatzoglou, A.; and Quadrana, M. 2016. The contextual turn: From context-aware to context-driven recommender systems. In *10th ACM Conference on Recommender Sys- tems*, 249–252.

Rendle, S. 2012. Factorization machines with libfm. *ACM Trans. Intell. Syst. Technol.* 3(3):57:1–57:22.

Shi, Y.; Larson, M.; and Hanjalic, A. 2014. Collaborative fil- tering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Comput. Surv.* 47(1):3:1– 3:45.

Zheng, Y.; Burke, R.; and Mobasher, B. 2013. Recommen- dation with differential context weighting. In *User Model- ing, Adaptation, and Personalization*. Springer. 152–164.

Zheng, Y.; Mobasher, B.; and Burke, R. 2014. Cslim: Con- textual slim recommendation algorithms. In *8th ACM Con- ference on Recommender Systems*, 301–304.

# Acknowledgements

Dr. Derek Bridge

Diego Carraro

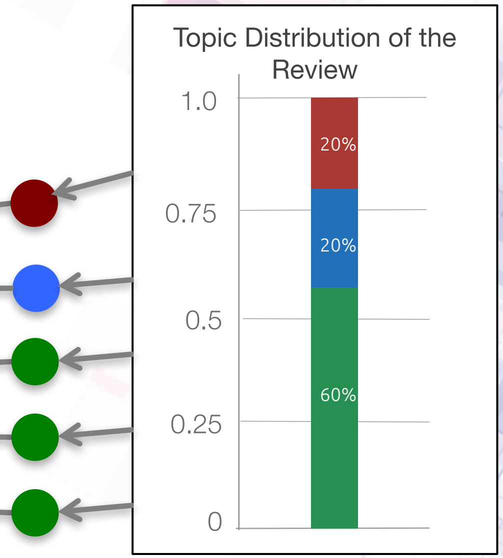
Mesut Kaya

Insight UCC

# Topic Modeling

- Each document is a random mixture of corpus-wide topics
- Each topic is composed of words that co-occur along documents

*"During the summer, we like to take a mini staycation. This year it was extra special as we also got engaged. Our stay at the Biltmore was just fantastic. The service exceptional, the food amazing- it was great at the pool, Wrights and also at Frank and Alberts. The only reason I am not giving it a full 5 stars is the 'upgraded' room was just a nice basic room. Though it was certainly nice, it wasnt what I expected for being the Biltmore. However, everything else certainly lived up to that expectation".*



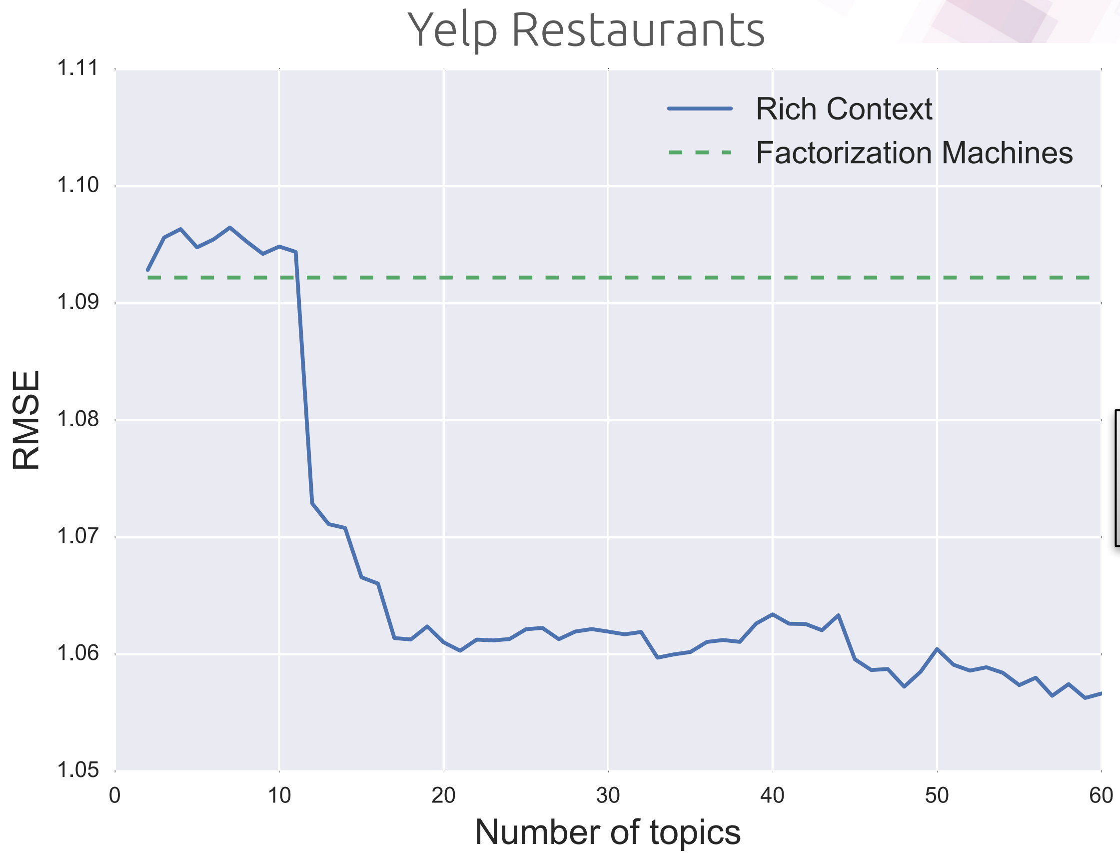

| | |
|---|---|
| Summer | (0.04) |
| Weekend | (0.02) |
| June | (0.01) |
| … | |

| | |
|---|---|
| Holiday | (0.05) |
| Romantic | (0.03) |
| Staycation | (0.01) |
| … | |

| | |
|---|---|
| Room | (0.05) |
| Pool | (0.04) |
| Sauna | (0.01) |
| … | |

| | |
|---|---|
| Free | (0.03) |
| Cheap | (0.02) |
| Expensive | (0.01) |
| … | |

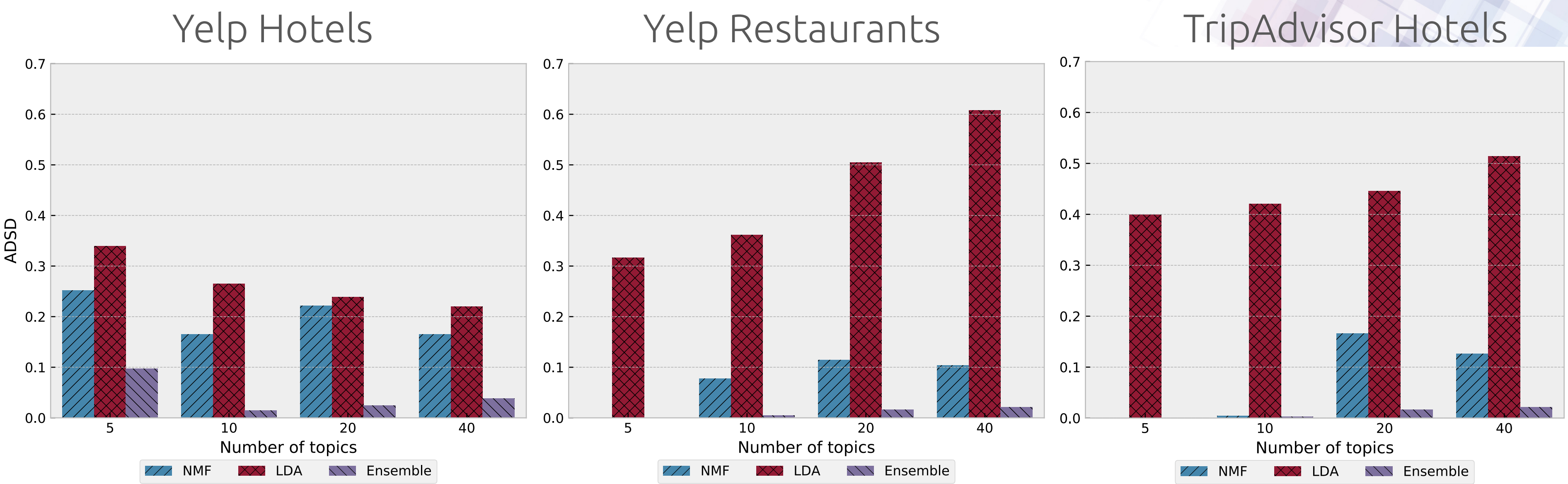
# Number Of Topics vs Performance




**The number of topics matters!**

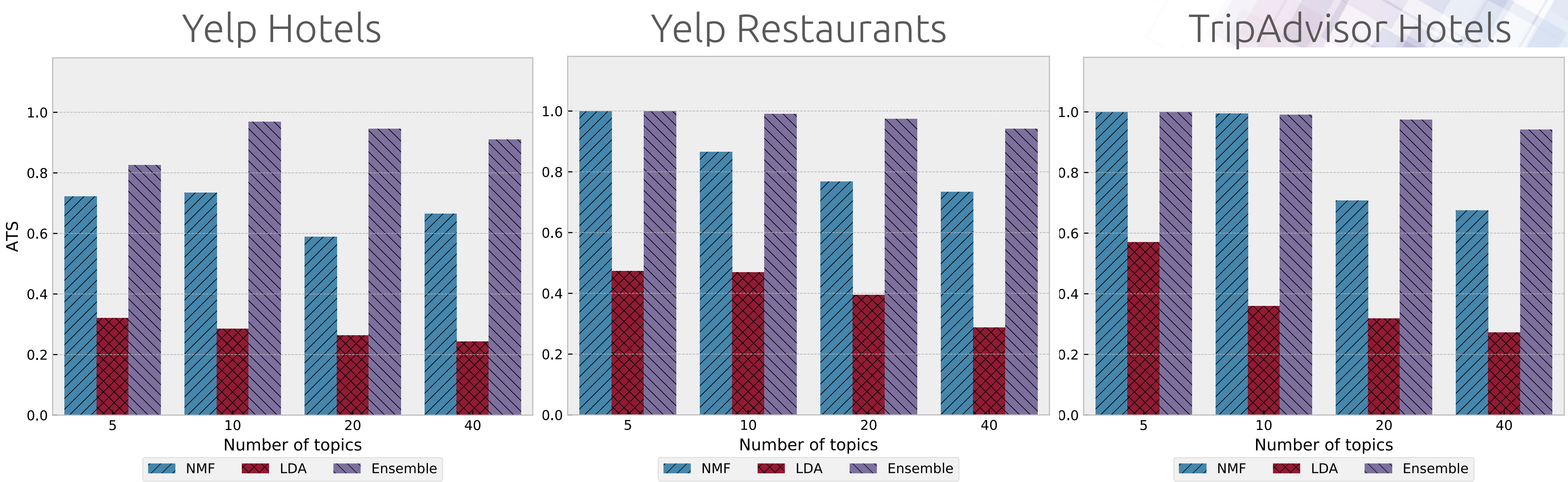
# Topic Model Validation (Stability)

**Average Descriptor Set Difference**



# Topic Model Validation (Stability)

**Average Term Stability**



# Topic Model Validation (Context-Richness)

| Topic 1 | | Topic 2 | | Topic 3 | | Topic 4 | |
|---|---|---|---|---|---|---|---|
| family | 0.194 | sushi | 0.271 | wife | 0.358 | service | 0.435 |
| sunday | 0.086 | bar | 0.055 | date | 0.026 | atmosphere | 0.023 |
| town | 0.069 | town | 0.021 | birthday | 0.020 | customer | 0.018 |
| brunch | 0.029 | spot | 0.020 | anniversary | 0.019 | table | 0.012 |
| weekend | 0.021 | saturday | 0.014 | weekend | 0.015 | drink | 0.011 |
| Score: 0.33 | | Score: 0.014 | | Score: 0.438 | | Score: 0.0 | |

Topic Model Score: 0.1955

$$ts(t) = \sum_{w} (p_{wt} * v_{wt})$$

The topic model score is the average of the topic scores

# Evaluation - Generated Topic Models (Yelp Restaurant)

| | Ratio | Word 1 | Word 2 | Word 3 | Word 4 | Word5 |
|---|---|---|---|---|---|---|
| **Topic 1** | 2.03 | night | dinner | friend | saturday | friday |
| **Topic 2** | 1.61 | lunch | today | day | friend | yesterday |
| **Topic 3** | 1.34 | time | couple | week | minute | hour |
| **Topic 4** | 1.1 | breakfast | morning | sunday | club | day |
| **Topic 5** | 1.07 | review | yelp | experience | star | read |
| **Topic 6** | 0.99 | scottsdale | location | town | experience | tempe |
| **Topic 7** | 0.93 | restaurant | phoenix | area | mexican | week |
| **Topic 8** | 0.85 | chicken | pizza | burger | sandwich | cheese |
| **Topic 9** | 0.75 | place | area | bar | love | home |
| **Topic 10** | 0.72 | food | service | mexican | atmosphere | price |

# Evaluation - Ranking Prediction

| Algorithm | Recall@10 | | |
|---|---|---|---|
| | Yelp | | TripAdvisor |
| | Hotels | Restaurants | Hotels |
| **Rich-Context** | **0.381** | **0.086** | **0.074** |
| Factorization Machines | 0.372 | 0.056 | 0.057 |
| BPMF | 0.150 | 0.014 | Out of memory |
| Biased MF | 0.068 | 0.016 | Out of memory |
| CAMF-C | 0.085 | 0.016 | Out of memory |
| CAMF-CI | 0.235 | 0.016 | Out of memory |
| CAMF-CU | 0.239 | 0.017 | Out of memory |
| CAMF-CUCI | 0.082 | 0.015 | Out of memory |
| DCR | 0.125 | Out of memory | Out of memory |
| DCW | 0.125 | Out of memory | Out of memory |
| **Improvement (all users)** | **2.32%** | **55.07%** | **30.15%** |

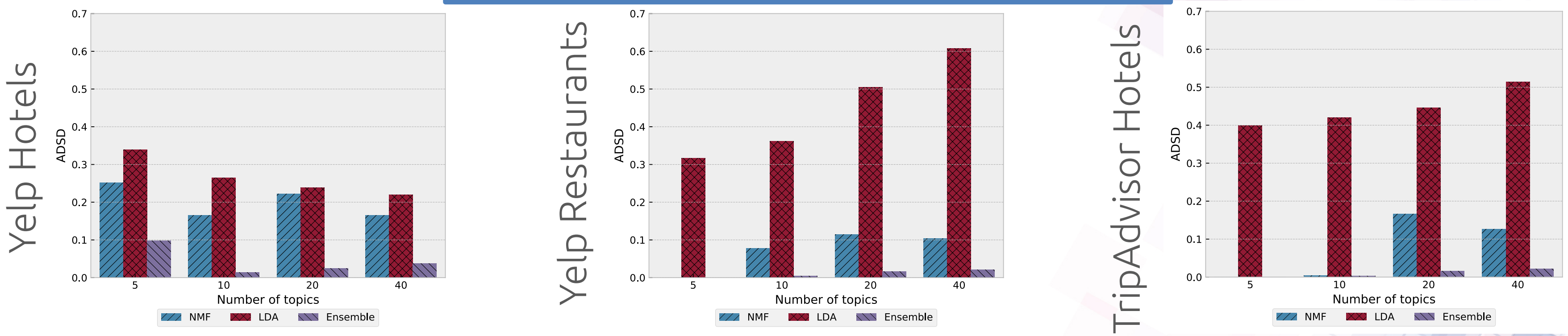
| | | | |
|---|---|---|---|
| **Improvement (new users)** | **3.67%** | **47.00%** | **20.50%** |

# Evaluation - Rating Prediction

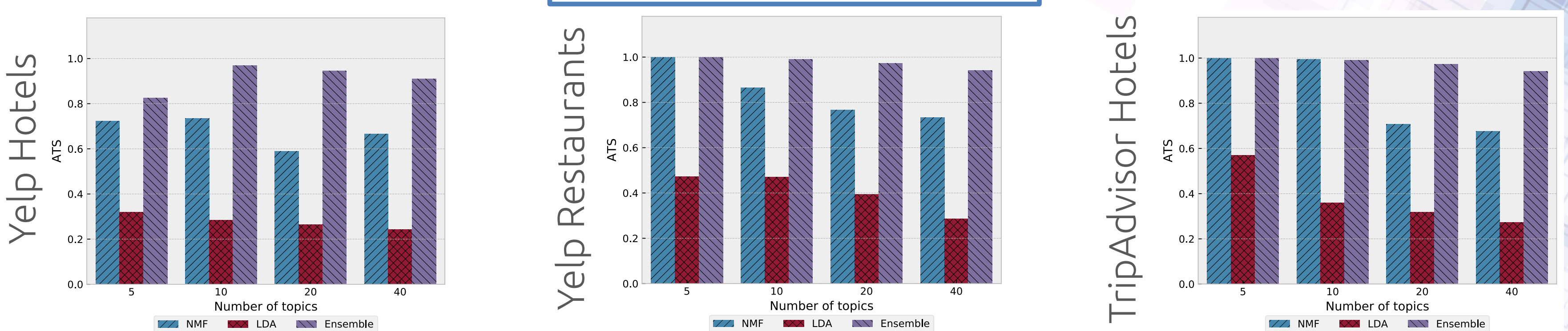
| Algorithm | RMSE | | |
|---|---|---|---|
| | Yelp | | TripAdvisor |
| | Hotels | Restaurants | Hotels |
| **Rich-Context** | **1.050** | **1.056** | **0.962** |
| Factorization Machines | 1.106 | 1.092 | 1.057 |
| BPMF | 1.495 | 1.405 | 1.435 |
| Biased MF | 1.109 | 1.183 | 1.104 |
| CAMF-C | 1.100 | 1.184 | 1.084 |
| CAMF-CI | 1.160 | 1.213 | 1.104 |
| CAMF-CU | 1.343 | 1.228 | 1.420 |
| CAMF-CUCI | 1.123 | 1.200 | 1.104 |
| DCR | 1.224 | Out of memory | Out of memory |
| DCW | 1.226 | 1.321 | Out of memory |
| **Improvement (all users)** | **4.45%** | **3.29%** | **8.99%** |
| **Improvement (new users)** | **5.43%** | **4.54%** | **9.96%** |

# Topic Model Validation (Stability)

## Average Descriptor Set Difference

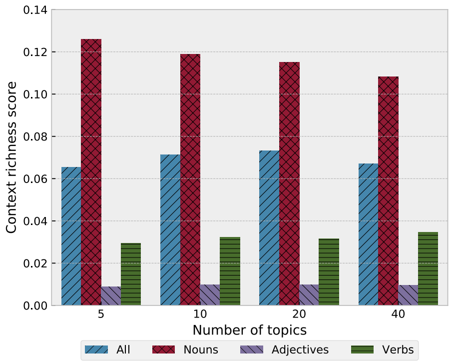


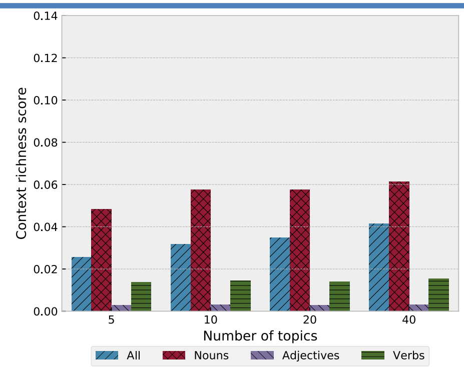

## Average Term Stability

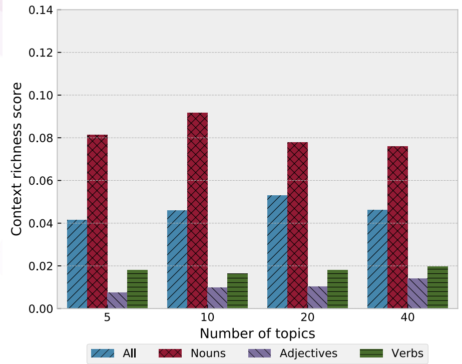


# Topic Model Validation (Context-Richness)

## Part of speech types

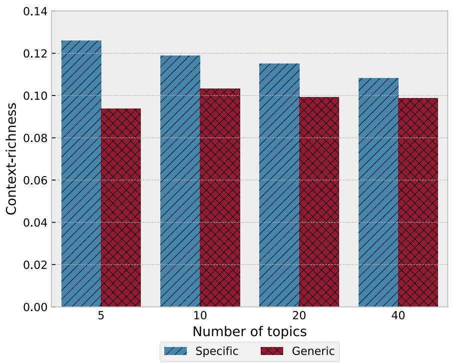


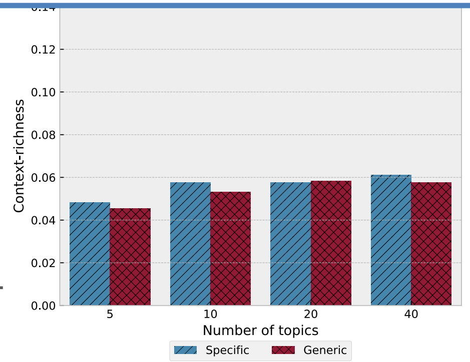




## Specific vs generic reviews

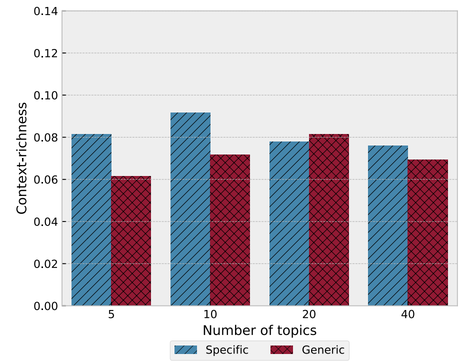
Yelp Hotels




Yelp Restaurants




TripAdvisor Hotels



## Evaluation - Ranking Prediction

Evaluation metric Recall@10

| 9 SOTA | | Rich-Context vs best SOTA | | |
|---|---|---|---|---|
| 6 CARS | 3 non-CARS | Yelp Hotels (+2.32%) | Yelp Restaurants (+55.07%) | TripAdvisor Hotels (+30.15%) |

# Contributions

- **Recommendations without predefining context**
  - New way of representing context.
- **Better prediction performance**
  - New methodology for offline ranking evaluation.
  - Better performance for brand-new users.
- **Extract context from reviews in an unsupervised way**
  - Methodology for selecting the best topic modeling algorithm.
  - New metrics to measure context-richness of topic models.
  - Improved methodology to classify reviews.